CN 53-1189/P ISSN 1672-7673

SExtractor及其在深度巡天研究中的应用*

阮志锋，邹文弢
（中国科学院云南天文台，云南 昆明 650011）

摘要：SExtractor作为一套从巡天图像中检测天体并提取天体星等、位置等信息的开源软件，已经被广泛运用在深度巡天的目标源提取和测量中。深度巡天中目标源的检测能力（深度）和对轮廓重叠的目标源的分解能力往往比光子噪声带来的测量误差更重要。针对不同研究目的的目标天体的提取往往需要在对目标源的检测能力和分解能力之间进行平衡，在使用SExtractor提取目标源时其参数设定也存在较大差异。介绍了SExtractor在深度巡天中参数的设定，结合分析讨论了SExtractor获取天体目标和测量的精度，同时也发现在深度巡天中使用SExtractor提取目标源存在十分严重的过度分化问题。

关键词：目标检测；深度巡天；光度学
中图分类号：TN216 **文献标识码**：A **文章编号**：1672-7673(2016)02-0266-07

SExtractor(Source-Extractor)是一套从巡天图像中检测天体并提取天体星等、位置等信息的开源软件[1]。它能快速从星系巡天图像中检测目标天体并测量天体的各项参数，而且对图像中轮廓(profile)部分重叠的天体具有较好的分辨和分解能力。与其它同类软件(如FOCAS[2]等)相比，SExtractor有处理速度快和输出参数灵活等优点，并保留了较高的测量精度。因此被广泛应用于天文巡天中的目标天体提取和天体的物理参数测量。本文对SExtractor软件的原理和参数设定做了系统的介绍，并结合分析讨论了SExtractor获取天体目标和测量的精度以及存在的问题。本文可作为未来国内较深巡天研究中应用SExtractor或类似软件的参考(如为郭守敬望远镜大科学项目观测准备目标源列表的巡天)，也可帮助国内研究者在使用由SExtractor检测的深度巡天数据时了解数据的来源和可能存在的问题。本文第1节介绍了SExtractor的原理和主要的参数设定；第2节介绍了SExtractor在具体深度巡天中的应用，并介绍了其精度问题；第3节进行了简单的总结和展望。

1 SExtractor软件的原理和参数设定

1.1 天光背景估计

通常一幅巡天图像中的每一个像素点的光子计数值都是背景信号和目标源信号的叠加。为了准确检测巡天中所有的目标源，需要对整个巡天天区构造天光背景分布。SExtractor在构建天光背景分布时采用K-σ Clipping法。具体的方法是将图像分解成若干个小的局部区域，在一个局部区域内计算其中所有像素的平均值(*mean*)、中值(*median*)和标准差(σ)，再去除那些落在$median \pm 3\sigma$之外的像素值，然后对剩下的像素值重复上述方法，如此反复多次(即迭代)直到剩下的所有像素值都在$mean \pm 3\sigma$之内。如果每次迭代后的σ相对于上一次迭代的σ变化都小于20%，对于此区域内的背景值直接用最终的平均值近似。如果σ改变大于20%，对于此区域内的背景值用最终的众数值(*mode*)近似。通常使用的天光背景直方图众数估计为①

$$mode = 3median - 2mean, \tag{1}$$

* 基金项目：中国科学院知识创新工程重要方向项目（KJCX2-YW-T24）资助.
收稿日期：2015-05-03；修订日期：2015-05-27
作者简介：阮志锋，男，硕士. 研究方向：星系统计物理. Email：ruanzhifeng@ynao.ac.cn

① Kendall M, Stuart A. The advanced theory of statistics. Vol.1: distribution theory

但在 SExtractor 的估计中文[1]使用了另外一个直方图众数值估计：

$$mode = 2.5median - 1.5mean \tag{2}$$

通过模拟表明，(2)式比(1)式更加准确，且结果受目标源密集度的影响相对较小，但会引入较大的噪声。

SExtractor 对所有小区域用 K-σ Clipping 法计算得到背景值随区域位置的分布，并对这个分布进行中值滤波平滑得到全图像区域的天光背景分布图像(参数 BACK_FILTERSIZE 为中值滤波时计算中值的像素范围)。这样有效地建立了可能不均匀的天光背景分布，也修正了一些亮星附近被过高估计的背景值。在 SExtractor 中每个小区域的面积(像素数量)等于输入参数 BACK_SIZE 的值，一个合适的 BACK_SIZE 大小对背景估计十分重要。如果不需要 SExtractor 扣除背景可以把参数 BACK_TYPE 设置为 MANUAL，同时把参数 BACK_VALUE 设置为 0.0，0.0。

1.2　目标源检测和提取

由于观测的天体没有统一固定的轮廓，对目标源的检测有两种经典的检测方法：峰值检测和阈值检测。峰值检测一般适用于恒星的检测，但不适用于低亮度面源(Low-Surface-Brightness，LSB)的检测[3]。SExtractor 中目标源的检测采用阈值检测法，由 DETECT_MINAREA、DETECT_THRESH 和 ANALYSIS_THRESH 3 个输入参数调节。参数 DETECT_THRESH 用于检测目标源的阈值，ANALYSIS_THRESH 是分析目标源 CLASS_STAR(恒星-星系分离参数)和 FWHM 时用到的阈值。一组像素被认定为一个天体轮廓而被提取需要满足 3 个条件：

(1)所有扣除背景后的像素值要高于设定的检测阈值；

(2)所有这些像素必须在一个连通区域内；

(3)这些像素的数量必须不少于设定的 DETECT_MINAREA 值(最小的像素个数)。

一幅 CCD 图像中除了包括目标源信号和天光背景，还存在一些噪声信号。SExtractor 通过对图像进行滤波以减弱噪声对目标源检测的影响。SExtractor 默认的卷积掩模为一个 3×3 的塔形矩阵，还提供了其它几个可供选择的卷积掩模，包括适用于暗弱天体检测的高斯掩模和适用于低亮度面源检测的 Top-hat 掩模等。

对于一个用阈值检测提取的连通像素区域(简称像素块)，可能是一个单一的目标源的轮廓，也可能包含了多个目标源轮廓的叠加。SExtractor 采取了一种多阈值算法构建一组多等光度线来分析判断一个像素块是否包含了多个目标源的叠加，并且对有叠加的情况分解其中不同目标源的像素区域。这个过程由输入参数 DEBLEND_NTHRESH 和 DEBLEND_MINCONT 调节。具体的流程是对每一像素块在检测阈值和块中扣除背景后的峰值之间分成 N 个水平(N 值即为 DEBLEND_NTHRESH 值)，这 N 个水平构成一线性或指数数列(t_i，$i=1, 2, \ldots N$)，相应地 SExtractor 在像素块内再作 N 次阈值检测，其中第 i 次以 t_i 为阈值，每次检测得到一组或多组连通的像素小区，以此构造一个块中目标源流强分布的“树”分布图(图 1)。对于第 i+1 次检测形成的某一分支，单独作为一个独立目标源的判断条件是：

(1)分支对应的连通小区 T_{i+1}的所有(扣除背景后的像素值大于 t_{i+1})像素的数量和总流强大于组合模块相应值的 DEBLEND_MINCONT 倍(一个比较合理的值为~0.005[1])。

(2)至少有另一个分支与(1)中的分支满足同一标准。这里组合模块指连接点处第 i 次检测形成的分支(或主干)对应的连通小区，即包含 T_{i+1}的 T_i。

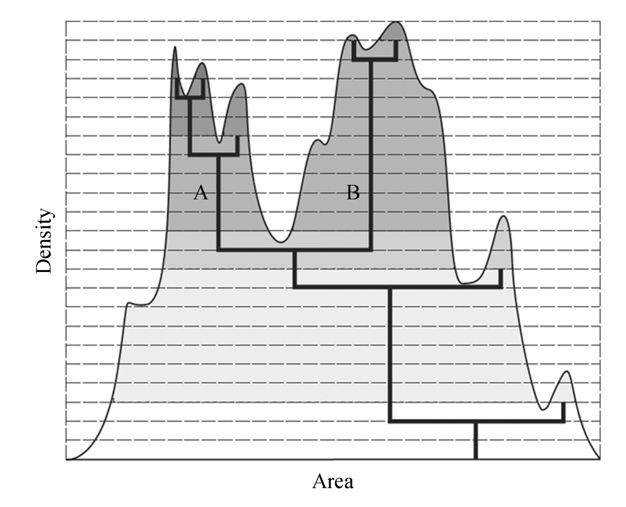


图 1　像素块流强分布的“树”分布图[1]。横线代表不同的分解检测阈值 t_i

Fig. 1　The "Tree structure" of the light distribution within the object

正常情况下，这种标准无法分离星等值相差约

6 个星等以上的目标源。对于图像边缘扣除背景后的像素值低于分离阈值的像素点，SExtractor 通过高斯函数拟合这些像素值计算像素点属于某一个子目标的概率，然后按概率将流强分配到对应的子目标天体中。

1.3 测量

SExtractor 的测量基于检测的目标源的流强分布，即在检测到的属于该目标源的像素区域内扣除天光背景后的像素值随位置的分布。如果参数 CLEAN 被设置为 Y，SExtractor 将用 Moffat 函数拟合与待测目标源相邻的其它源的流强分布，并把拟合所得的 Moffat 轮廓从待测目标源流强分布中扣除。Moffat 轮廓满足方程：

$$\frac{I(r)}{I(0)}=\frac{1}{(1+k\times r^2)^{\beta}} . \tag{3}$$

式中，β 为参数 CLEAN_PARAM，其大小应取在 0.1 到 10 之间。最基本的天体数据测量是确定目标源的中心。SExtractor 使用修正矩[4]方法对一个目标源求中心，其定义如下：

$$X=\bar{x}=\frac{\sum I_i x_i}{\sum I_i} , \tag{4}$$

$$Y=\bar{y}=\frac{\sum I_i y_i}{\sum I_i} . \tag{5}$$

上式中的求和对经阈值检测和像素块分割后所有属于一目标源的像素点进行(见上一节)。(x_i、y_i)为目标源中一像素点在图像中的位置；I_i 为这个像素点扣除背景和分解后的像素值(流强)。可见中心的位置受检测阈值和重叠源分离设置的影响。为了对巡天中各种目标源实现有效的测光，SExtractor 提供了包括孔径测光在内的 4 种测光方式。

(1)等光度(ISO)测光将经检测和分解后的所有归于待测目标源的像素点(见上一节)的像素值求和估计流强，得到目标源的光子计数并换算为流强及星等。等光度测光的精度受到检测阈值(参数 DETECT_THRESH)和背景估计的影响较大。

(2)圆孔孔径(APERTUPE)测光是用户自定义一组孔径值(参数 PHOT_APERTURES)对目标源落在孔径内的像素值 I_i 求和。圆孔测光的精度主要取决于选择的圆孔大小，过大或者过小的孔径都会对目标源的光度测量带来较大的误差。

(3)修正等光度(ISOCORR)测光是对等光度测光方法的改进，考虑到在等光度测光中常常会遗漏目标源的一些低于检测阈值的流强，在修正中 SExtractor 以二维对称高斯函数作为目标天体流强分布的一级近似。在这个近似下检测阈值对应的等光度线包围的区域内的流强占目标源总流强的比例 $\eta=I_{iso}/I_{tot}$ 满足如下关系：

$$1-\frac{1}{\eta}\ln(1-\eta)=\frac{At}{I_{iso}} . \tag{6}$$

式中，A 是检测阈值对应的等光度线所包围的面积；t 是检测阈值。上式没有解析解，但是可以得到它的一个近似级数展开解：

$$\eta\approx 1-0.1961\frac{At}{A_{iso}}-0.7512\left(\frac{At}{I_{iso}}\right)^2 . \tag{7}$$

当 $\eta>0.4$ 时，(7)式的误差小于 10^{-2}。星体总的星等值 m_{tot} 是对等光度测光星等的修正：

$$m_{tot}=m_{iso}+2.5\log(\eta) . \tag{8}$$

这种方法对恒星的测光有较高的精度，对轮廓接近高斯分布的展源也适用。

(4)自动孔径(AUTO)测光在待测目标源周围使用一个按具体目标源调整的椭圆孔径(即 Kron 孔径[5])，对孔径内的目标源的流强求和得星等。目标源流强分布的对称一阶矩为

$$r_1=\frac{\sum rI(r)}{\sum I(r)} . \tag{9}$$

式中，$I(r)$ 为距中心 r 处的 I_i 的平均。文[6]证实了 Kron 孔径作为一个大小灵活的孔径能够将目标大

部分的流强包含在孔径内，有超过 90%的流强落在半径为 $2r_1$ 的圆孔孔径内。所以在长短半轴分别为 ϵkr_1 和 kr_1/ϵ 的一个椭圆区域内测光时，$k\approx2$ 是平衡系统误差与随机误差比较好的选择。对于信噪比较低的情况，为了避免孔径过小引起过大的随机误差，SExtractor 允许在测光时限定一个最小的孔径 R_{min}。其中 k 和 R_{min} 的大小由 PHOT_AUTOPARAMS 参数给定。一般情况下用 Kron 孔径能够较精准地得到目标源的流强值，但是在目标源附近有比其亮得多的源时，Kron 孔径测光会引入较大的误差而使测光精度降低，此时修正等光度测光比 Kron 孔径测光更精准。

天体的星等(m)由公式 $m=-2.5\times\log R_{10}R(I)+Z_p$ 给出，其中 I 为测得的目标源的总流强；Z_p 为星等零点(参数 MAG_ZEROPOINT)。星等的泊松误差为 $\Delta m=1.0857\sqrt{A\sigma^2+I/g}/I$，其中 A 为用于计算流强的面积(像素数)；σ 是通过背景估计得到的背景噪声与读取误差的合成涨落的标准差；g 为探测器的增益(等于输入参数增益)。图 2 为对模拟的孤立星系用不同方法测光时平均遗漏流强随星系星等的变化。针对前述对待测源 Kron 星等受附近亮源的影响，SExtractor 也提供了自动选择 Kron 星等和修正等光度星等的最佳星等 MAG_BEST 值。大多数情况下 MAG_BEST=MAG_AUTO(Kron 星等)，但当被扣除的周围源的流强高于所测目标源的总流强(见(3)式)的 10%时，MAG_BEST=MAG_ISOCORR(修正等光度星等)。

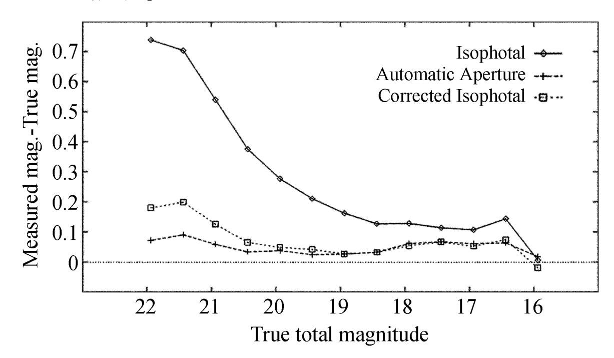


图 2　不同方法测光中平均遗漏的流强随目标源星等的变化②

Fig. 2　Flux lost with different faint-object photometry techniques as a function of total magnitude

2　SExtractor 在深度巡天中的应用

经验研究者一般认为 Kron 星等(MAG_AUTO)能较准确地反映星系的总流强。但在获取多波段颜色数据时，通常选择相同的孔径大小(常为 3″)测量星系的星等值，因为在星系颜色的测量中孔径遗漏的星系轮廓部分流强对结果的影响往往比星系总星等的误差造成的影响要严重得多。

在深度巡天中关注如何挑选那些较小较暗弱的目标源以达到足够的检测深度，所以在使用 SExtractor 检测目标源时应采用较低的阈值(对应 DETECT_THRESH)。但使用较低的阈值时，SExtractor 检测不可避免地出现一些问题：(1)过度分化那些面亮度较低的旋涡星系或不规则星系；(2)过度分解一些亮的展源；(3)错误地把一些亮星的散射光当成真正的源；(4)不能很好地分离小间隔的星系对。

针对这些问题文[7]提出了一个“冷-热”检测的方法，文[8]在 COSMOS 巡天的目标源检测中也使用了这一方法。他们在提取目标源时分别使用两组不同的参数设置运行 SExtractor，一组参数用于检测较大较亮的目标源(“冷”检测)，以保证亮源不会被过度分化，并尽可能抑制因为亮星的散射光而产生的虚假源。另一组参数则用于检测和分离较暗的目标源(“热”检测)，以保证检测有足够的深度。这种两步检测的方法还允许根据目标源典型的大小调整 BACK_SIZE 的值，以提高检测能力和测量精度。对“热”检测中得到的源列表在剔除那些落在“冷”检测所得的源周围 0.6″范围内的源之后，将其与“冷”检测的列表合并从而得到最终的源列表。具体的参数选择见表 1。

分析了文[8]使用“冷-热”检测方法得到的 COSMOS ACS 图像的目标源列表，并选用文[9]在 2009 年的一个用于星系形态学研究的列表作为对比。文[9]的列表是对文[8]列表中的星系进行更为细致的形态学分类，但在具体分类时剔除了列表中缺乏明确的形态学类别的源。通过匹配发现在文[8]的星系列表中大约有 2.4%的星系周围 1″范围内有因为没有具体的形态分类而在文[9]的列表中被舍弃的源，通过抽查部分图像发现绝大多数为可能被过度分化的星系。此外，抽查了部分成员相距

② Bertin E, Arnouts S. SExtractor: source extractor

很近的星系对(投影距离小于~10 kpc)的 ACS 图像，并进行了目视审查发现有相当一部分星系对不能被很好确定是被过度分化的星系还是真实的物理星系对。图 3 展示了抽查中发现的 9 个典型的可能被过度分化的星系的 ACS 图像。图像中的两个星系可以被解释为一个星系的两部分，即一个星系核和一个大规模恒星形成区域。这支持了文[8]对 COSMOS 数据中存在过度分化的结论。未来需要对这些星系进行进一步研究以确定是否被 SExtractor 过度分化。如确实存在过度分化，需探讨对 SExtractor 目标源检测结果进行修正的方式。

表 1 Leauthaud 等使用的 SExtractor 参数设置[8]

Table 1 SExtractor parameters setting by Leauthaud et al. (2007)

参数	亮源	暗源
DETECT MINAREA	140	18
DETECT THRESH	2. 2σ	1. 0σ
DEBLEND NTHRESH	64	64
DEBLEND MINCONT	0. 04	0. 065
CLEAN PARAM	1. 0	1. 0
BACK SIZE	400	100
BACK FILTERSIZE	5	3

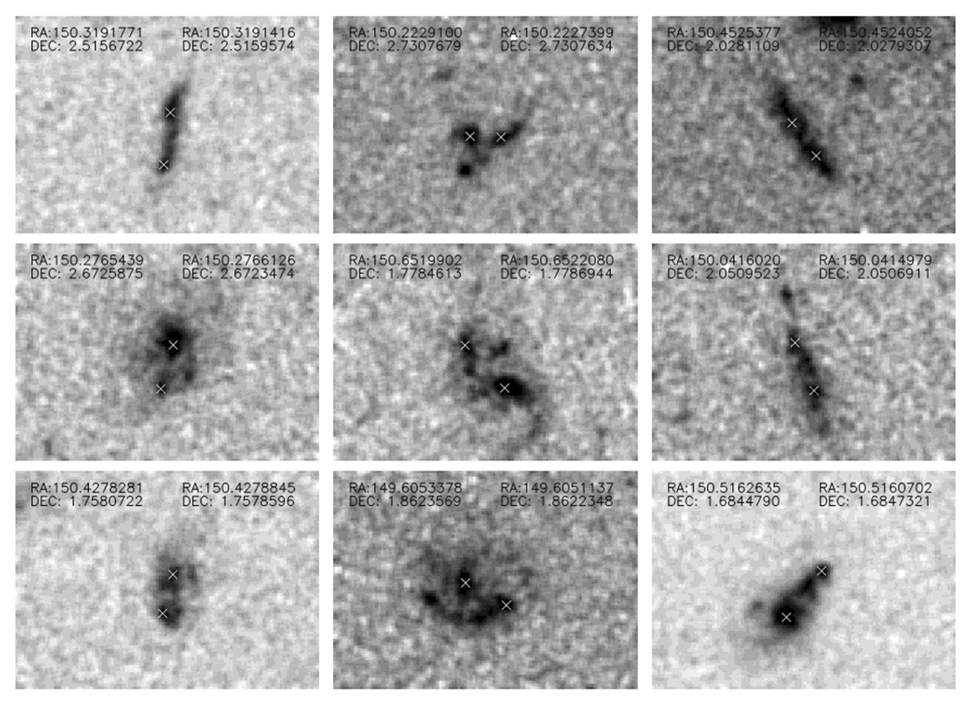


图 3 文[10]2007 年的星系列表中 9 个典型的可能被过度分化的星系的 ACS 图像

Fig. 3 Nine typical ACS galaxy images in the list of Leauthaud et al. (2007) which might be overdeblended

在设置 SExtractor 参数时，一个比较好的方法是文[10]在 CFH12K-VIRMOS 深度巡天的目标源检测中使用的重新取回人工模拟星系的方法。在参数设定之前先在图像中加入大量的人工模拟的星系，这些星系的形态和光度分布由基于现有的一些星系结构和演化模型的模拟得到，其位置和倾角在图像中随机分布；然后调整参数使得 SExtractor 目标源检测能重新找回最大数量的人工模拟星系。一般情况下地面图像和处理结果由于受大气扰动的影响分辨率较低，视间距较小的目标源轮廓的重叠情况要比空间巡天数据严重得多，所以应选择较小的 DEBLEND_MINCONT 以保证轮廓重叠的目标源能有较好的分离。例如，在 CFHTLS 深度巡天③中选取的 DEBLEND_MINCONT = 0. 002，在 COSMOS 地面巡天数据[11]中则选取更为极端的分离标准 DEBLEND_MINCONT = 0. 0。此外，空间巡天测光由于不受大气的影响对地面数据有较好的校准作用。在地面巡天区域如果包含有对应的空间望远镜的巡天数据，也可以通过对比地面数据中检测的目标源与对应的空间数据，优化地面数据中目标源提取和测量的 SExtractor 参数。

图 4 是对 HDF-N 天区[2]内部分表现为孤立源的地面和空间图像测光对比。图中实线为地面数据

③ Cuillandre J C J, Withington K, Hudelot P, et al. Introduction to the CFHT Legacy Survey final release (CFHTLS T0007)

测光(Hawaii HDF-N 巡天数据[12])的光度增长曲线，虚线为空间数据测光的光度增长曲线。R 为离源中心位置的角距离(单位为角秒)，对应的星等值为半径 R 的圆孔孔径星等。交叉号和三角形分别为地面数据测光和空间数据测光中的 Kron 孔径星等，被置于 $R=1.5''$ 处(对应直径为 3″的圆孔孔径)以与 3″孔径星等对比。从图中可发现 Kron 星等与总星等(即增长曲线所趋平时的值)基本符合，3″孔径星等也较为准确，误差大多<0.3 mag。这显示了通常使用的 SExtractor 的 Kron 星等一般是十分准确的测光方式，3″孔径星等也较为准确。

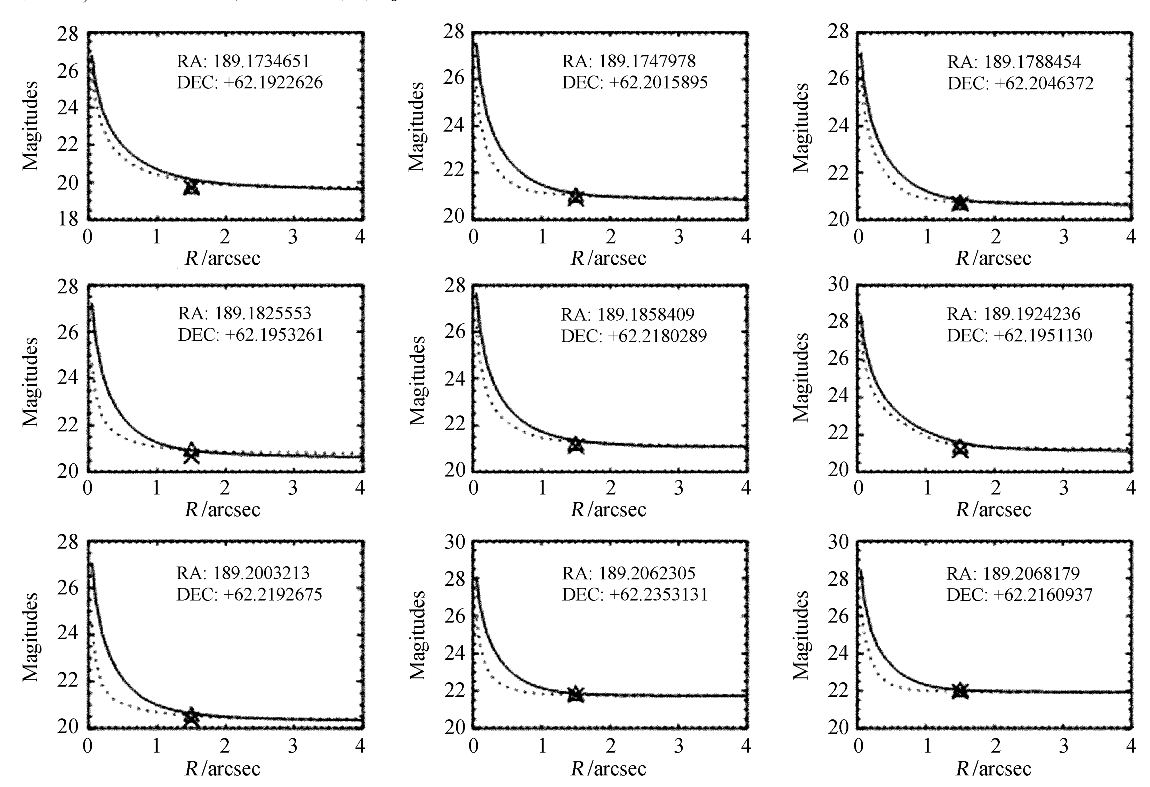


图 4　从 HDF-N 中抽取的典型的表现为孤立星系的地面数据测光与空间数据测光的对比

Fig. 4　Comparing the isolated galaxy photometry between space-based and ground-based observations

3　总结与展望

SExtractor 以其处理速度快和输出参数灵活等优点在深度巡天中得到了良好的应用和发展。对存在的一些问题，比如检测列表的完整性以及大量因噪声而导致的假源的清除等已经有了较好的解决方法，然而对于一些星系可能的过度分化仍然是一个难题。特别是对于高分辨率的巡天图像，现有的检测算法可能得到改善。随着多项高分辨率多波段巡天工作的开展，一个可能的解决方案是将颜色和形态信息用于目标源的检测和识别过程，以求更准确地检测和分离目标源。

参考文献：

[1]　Bertin E, Arnouts S. SExtractor: software for source extraction [J]. Astronomy & Astrophysics Supplement Series, 1996, 177: 393-404.

[2]　Williams R E, Blacker B, Dickinson M, et al. The Hubble Deep Field: observations, data reduction, and galaxy photometry [J]. The Astronomical Journal, 1996, 112(4): 1335-1752.

[3]　Yee H K C. A faint-galaxy photometry and image-analysis system [J]. Publications of the Astronomical Society of the Pacific. 1991, 103(662): 396-411.

[4]　Stone R C. A comparison of digital centering algorithms [J]. The Astronomical Journal, 1989,

97: 1227-1237.

[5] Kron R G. Photometry of a complete sample of faint galaxies [J]. The Astrophysical Journal Supplement Series, 1980, 43: 305-325.

[6] Graham A W, Driver S P. A concise reference to (projected) Sérsic R1/n quantities, including concentration, profile slopes, petrosian indices, and Kron magnitudes [J]. Publications of the Astronomical Society of Australia, 2005, 22(2): 118-127.

[7] Rix H W, Barden M, Beckwith S V W, et al. GEMS: galaxy evolution from morphologies and SEDs [J]. The Astrophysical Journal Supplement Series, 2004, 152: 163-173.

[8] Leauthaud A, Massey R, Kneib J P, et al. Weak gravitational lensing with COSMOS: galaxy selection and shape measurements [J]. The astrophysical Journal Supplement Series, 2007, 172: 219-238.

[9] Tasca L A M, Kneib J P, Iovino A, et al. The zCOSMOS redshift survey: the role of environment and stellar mass in shaping the rise of the morphology-density relation from z~1 [J]. Astronomy & Astrophysics, 2009, 503(2): 379-398.

[10] McCracken H J, Radovich M, Bertin E, et al. The VIRMOS deep imaging survey. II: CFH12K BVRI optical data for the 0226-04 deep field [J]. Astronomy & Astrophysics, 2003, 410: 17-32.

[11] Capak P, Aussel H, Ajiki M, et al. The first release COSMOS optical and near-IR data and catalog [J]. The Astrophysical Journal Supplement Series, 2007, 172: 99-116.

[12] Capak P, Cowie L L, Hu E M, et al. A deep wide-field, optical, and near-infrared catalog of a large area around the Hubble Deep Field North [J]. The Astronomical Journal, 2004, 127: 180-198.

The SExtractor and Its Applications in Deep Sky Surveys

Ruan Zhifeng, Wu Wentao

(Yunnan Observatories, Chinese Academy of Sciences, Kunming 650011, China, Email: ruanzhifeng@ynao.ac.cn)

Abstract: Deep sky surveys provide fundamental data for studies of formation and evolution of galaxies, galaxy clusters, large-scale structures of the universe, and supermassive black holes. The SExtractor software package is an open-source software system. It can build a catalogue of objects from a sky-survey image and output magnitudes, surface coordinates, sizes, shape descriptions, and other information of objects. The SExtractor has been widely used for detection and measurement of objects in deep sky surveys. For object detection and measurement, abilities to detect objects (for completeness) and to separate objects of superimposed profiles (for resolving confusion) often have a larger impact on results than measurement errors due to photon noise. This requires certain balance between the abilities to detect objects and to separate objects of superimposed profiles. When using the SExtractor, the parameter values should then vary according to different types of objects so as to optimize detection and measurement. To detect apparently small faint objects, low detection thresholds need to be adopted, but low detection thresholds can lead to overdeblending of high-redshift galaxies of complex inner structures, or confusion of outer features of bright galaxies/scattered-light features of bright stars to be independent objects. Particularly, overdeblending can cause severe contamination in statistics of high-redshift close pairs identified from deep sky surveys. Confusion from parts of bright galaxies and scattered-light features of stars can be avoided by flexible settings of SExtractor parameter values. Overdeblending of faint galaxies of complex structures remains a difficult problem though. In this paper, we systematically describe the principles and parameters of the SExtractor. By incorporating our own analysis we also discuss accuracies and other measurement problems in using the SExtractor. At last, we give some suggestions to those problems.

Key words: Object detection and measurement; Deep sky survey; Photometry